

QÜESTIÓ, vol. 25, 3, p. 437-477, 2001

ESTIMACIÓN NO PARAMÉTRICA DE LA FUNCIÓN DE RIESGO: APLICACIONES A SISMOLOGÍA

GRACIELA ESTÉVEZ PÉREZ
ALEJANDRO QUINTELA DEL RÍO
Universidad de A Coruña*

Se estudia la estimación de tipo no paramétrico de la función de riesgo o razón de fallo de una variable aleatoria real. A partir de una muestra X_1, X_2, \dots, X_n de datos no censurados y no necesariamente independientes, se considera un estimador cociente entre el estimador núcleo de la función de densidad y un estimador núcleo de la función de supervivencia, sobre el que se estudia el problema de selección del parámetro ventana. Por medio de un estudio de simulación se observa la ventaja de utilizar este estimador frente al que estima la función de supervivencia a través de la función de distribución empírica. Finalmente, se realiza una aplicación práctica a datos de terremotos sucedidos en California y en Granada.

Palabras clave: Estimación no paramétrica, procesos fuertemente mixing, validación cruzada, razón de fallo

Clasificación AMS (MSC 2000): 62G05, 62G20, 62M99

*Departamento de Matemáticas-Facultad de Informática. Universidad de A Coruña. Campus de Elviña, s/n. 15071 A Coruña. Teléfono: 981167000, ext.: 2109.

–Recibido en julio de 1999.

–Aceptado en octubre de 2001.

1. INTRODUCCIÓN

En análisis de supervivencia, el interés se centra, por un lado, en una población homogénea en la que se considera un suceso puntual, normalmente llamado *fallo*; y, a la vez, en el tiempo de ocurrencia de dicho suceso, generalmente llamado *tiempo de fallo*. Para determinar el *tiempo de fallo* de forma precisa debe definirse sin ambigüedad el tiempo origen, la escala de medida de ese tiempo, y el significado del suceso puntual *fallo*. Tal suceso puede ser, por ejemplo, la muerte de pacientes tras algún tipo de trasplante, la avería de algún componente de una máquina, la ruptura de algún material sometido a una fuerza o la ocurrencia de un terremoto. De esta forma, el tiempo de fallo será, respectivamente, el tiempo de vida después del trasplante, el tiempo de duración de una máquina hasta su primera avería, el tiempo de ruptura del material y el tiempo entre terremotos sucesivos. Esto nos muestra como el ámbito de aplicación del análisis de supervivencia abarca disciplinas de índole tan diversa como la medicina, la química, estudios de fiabilidad o sismología.

Supongamos que X es una variable aleatoria real, continua, con función de densidad $f(\cdot)$ y función de distribución $F(\cdot)$. Además de estas formas matemáticamente equivalentes de representar la distribución de X , existen otras funciones particularmente útiles en análisis de supervivencia, como son la función de supervivencia o la función de razón de fallo.

La función de supervivencia se representa por $S(x)$ y se define como la probabilidad de que el tiempo de fallo sea superior a x , es decir,

$$S(x) = P(X > x).$$

De la definición de función de distribución $F(\cdot)$ y de su relación con la función de densidad $f(\cdot)$ se tiene que $f(x) = (-dS(x))/dx$. En aplicaciones médicas, la función de supervivencia, como su propio nombre indica, se interpreta como la probabilidad de que un individuo en estudio sobreviva al tiempo x , y ha sido ampliamente estudiada en contextos de censura y/o truncamiento (ver por ejemplo, Lee (1992) o Klein y Moeschberger (1997)).

La función de razón de fallo, que también es conocida como «función de azar», «fuerza de mortalidad», «función de intensidad» o «**función de riesgo**» se representa por $r(\cdot)$ y se define como

$$(1) \quad r(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X < x + \Delta x \mid X \geq x)}{\Delta x}.$$

Por la definición de probabilidad condicionada tenemos que

$$(2) \quad r(x) = \frac{f(x)}{S(x)} = \frac{f(x)}{1 - F(x)}.$$

Si la variable X mide el tiempo de fallo, la función $r(\cdot)$ es una medida de la predisposición de fallo como una función del tiempo, en el sentido en que $r(x)\Delta x$ representa la proporción esperada de individuos que, sobreviviendo al tiempo x , fallan en el intervalo $(x, x + \Delta x)$. De esta forma, puede ser interpretada como el riesgo instantáneo de fallo en cada instante x sabiendo que el último fallo se ha producido en el instante 0.

En muchas situaciones prácticas es evidente que el conocimiento de la función de razón de fallo de una variable en estudio supondría grandes ventajas, de ahí el interés de estimar dicha función. Para tal propósito, el primer planteamiento que surge es considerar una familia paramétrica de distribuciones (Weibull, Gamma o Exponencial son algunos de los modelos más utilizados en casos prácticos) y, utilizando un conjunto de observaciones X_1, \dots, X_n de la variable X , estimar el vector de parámetros desconocido. En Cox y Oakes (1984) aparece una descripción de algunas de las familias paramétricas de distribuciones más usadas, así como criterios para la elección adecuada de tales familias. También, en su capítulo 3 puede verse una revisión de métodos de inferencia paramétrica basados en la función de verosimilitud. Otras obras de consulta sobre este tema son los anteriormente citados Lee (1992) y Klein y Moeschberger (1997).

Sin embargo, en muchos casos no se dispone de información suficiente para precisar la familia de distribuciones a la que pertenece la variable en estudio. En tales situaciones es deseable disponer de métodos no paramétricos para estimar la razón de fallo a partir de la información que proporciona un simple conjunto de observaciones X_1, \dots, X_n .

El problema de estimación no paramétrica de la función de razón de fallo ha sido tratado con bastante profundidad en la literatura, bajo la hipótesis de datos independientes. Algunas referencias al respecto son Watson y Leadbetter (1964a y b), Ahmad (1976), Hollander y Proschan (1984), Prakasa Rao y Van Ryzin (1985), Singpurwalla y Wong (1983) y Hassani, Sarda y Vieu (1986).

En ciertas ocasiones, sin embargo, la hipótesis de independencia entre las observaciones X_1, \dots, X_n no tiene por qué ser adecuada. En estudios sismológicos, por ejemplo, es mucho más realista suponer que los tiempos entre terremotos son dependientes, aunque esta dependencia se vaya debilitando para movimientos separados temporalmente. Esto nos llevará a considerar que las observaciones pertenecen a una serie de tiempo —que supondremos estrictamente estacionaria— y que cumplirá alguna condición de dependencia que sea razonable desde un punto de vista práctico, y que permita una manipulación matemática sencilla.

En este trabajo consideraremos la estimación no paramétrica de la función de razón de fallo (2) bajo condiciones de dependencia entre los datos y no censura. En la sección 2 definimos el estimador utilizado y las condiciones matemáticas de dependencia. Detallamos también algunas propiedades matemáticas del mismo, junto con una manera de elegir el parámetro de suavización en los estimadores no paramétricos. En la sección 3 se exponen algunos ejemplos de simulación que evidencian la mejora de utilizar el estimador definido en la sección 1 frente a un estimador utilizado en Sarda y Vieu (1989). En la sección 4 se analiza el procedimiento de selección del parámetro ventana desde un punto de vista local. Finalmente, en la sección 5 se exponen dos ejemplos de utilización práctica de la estimación no paramétrica del riesgo instantáneo de ocurrencia de terremotos, en dos regiones geográficas y periodos de tiempo diferentes.

2. ESTIMACIÓN NÚCLEO DE LA FUNCIÓN DE RIESGO BAJO CONDICIONES DE DEPENDENCIA

Muchos de los estimadores de la función de riesgo o razón de fallo y, en particular, el que consideraremos en este trabajo, se definen como un cociente entre un estimador no paramétrico de la función de densidad $f(\cdot)$ y un estimador no paramétrico de la función de supervivencia $1 - F(\cdot)$ (Watson y Leadbetter (1964a y b)). Nosotros consideraremos el estimador obtenido a partir de una muestra X_1, \dots, X_n dado por

$$(3) \quad r_h(x) = \frac{f_h(x)}{1 - F_h(x)},$$

donde $f_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$ es el conocido estimador núcleo de Parzen-Rosenblatt, y $F_h(\cdot)$ es el estimador núcleo de $F(\cdot)$, definido por

$$(4) \quad F_h(x) = \frac{1}{n} \sum_{i=1}^n H\left(\frac{x - X_i}{h}\right) = \int_{-\infty}^x f_h(t) dt.$$

La función núcleo (kernel) $K: \mathbb{R} \rightarrow \mathbb{R}$ verifica $\int K(u) du = 1$ y $H(\cdot)$ se obtiene a partir de $K(\cdot)$ mediante $H(x) = \int_{-\infty}^x K(u) du$. El parámetro $h = h(n) \in \mathbb{R}^+$ es el llamado *parámetro de suavización* o *ventana*, que controla el grado de suavización inherente en todo proceso de estimación de tipo no paramétrico.

Este estimador presenta varias ventajas con respecto al estimador que utiliza la función de distribución empírica para estimar la función $F(\cdot)$:

$$(5) \quad r_{h,1}(x) = \frac{f_h(x)}{1 - F_n(x)}, \text{ con } F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x\}}.$$

En primer lugar, este último estimador no es continuo, puesto que $F_n(\cdot)$ tampoco lo es. Además, $F_n(\cdot)$ presenta una deficiencia relativa con respecto a $F_h(\cdot)$ en términos de error cuadrático medio (Reiss (1981)). Un ejemplo de esta deficiencia se pondrá de manifiesto en la sección 3 en simulaciones. Este estimador ha sido estudiado por Sarda y Vieu (1989) en distintos contextos de dependencia (entre ellos el de fuertemente mixing, definido más adelante). Dichos autores demuestran propiedades de consistencia fuerte casi segura para el estimador, y estudian las propiedades de optimalidad de los procedimientos de validación cruzada global y local para la selección del parámetro ventana.

El estimador dado en (3) ha sido estudiado por Youndjé, Sarda y Vieu (1996) en un contexto de independencia. Para poder trabajar bajo la hipótesis de que los datos de la muestra en estudio sean dependientes, consideramos la siguiente definición:

Definición 2.1. Sea $\{X_t; t \in \mathbb{N}\}$ una sucesión de variables aleatorias, y sea \mathcal{F}_k^{k+n} la σ -álgebra generada por las variables $X_k, X_{k+1}, \dots, X_{k+n}$. Definamos

$$\alpha(n) = \sup_{k \in \mathbb{N}} \left\{ |P(A \cap B) - P(A)P(B)|; A \in \mathcal{F}_0^k, B \in \mathcal{F}_{k+n}^\infty \right\}$$

Entonces, se dice que $\{X_t; t \in \mathbb{N}\}$ es fuertemente mixing o α -mixing si $\lim_{n \rightarrow \infty} \alpha(n) = 0$.

Esta condición es de las más débiles entre las utilizadas en estudios con datos dependientes. La monografía de Doukhan (1994) es un buen referente para el estudio de éstas y otras condiciones similares. En dicha monografía se demuestra que un gran número de procesos dependientes cumplen condiciones de tipo mixing. Por ejemplo, los procesos ARMA estacionarios con ruido absolutamente continuo son fuertemente mixing. Para más referencias sobre estimación de tipo núcleo bajo hipótesis de datos mixing consultar, por ejemplo, Györfi *et al.* (1990).

Para poder trabajar en la práctica con $r_h(\cdot)$ es necesario elegir una función núcleo $K(\cdot)$, así como la ventana $h = h(n)$. Mientras la elección de $K(\cdot)$ no resulta demasiado importante, la elección de la ventana h juega un papel preponderante en el comportamiento del estimador. En Estévez y Quintela (1999) se estudian propiedades de consistencia puntual y uniforme de tipo casi seguro para el estimador $r_h(\cdot)$ bajo la suposición de que la muestra X_1, \dots, X_n es fuertemente mixing. Se prueba también la normalidad

asintótica del mismo y se describe un método de validación cruzada para la selección del parámetro ventana h , del que se prueba su optimalidad asintótica.

Una ventana es considerada óptima si minimiza algún criterio de error para el estimador $r_h(\cdot)$, como por ejemplo el *Error cuadrático promedio*:

$$(6) \quad ASE(h) = \frac{1}{n} \sum_{i=1}^n (r_h(X_i) - r(X_i))^2 W(X_i),$$

donde $W(\cdot)$ es alguna función de pesos no negativa, introducida para evitar el llamado «efecto frontera», propio de los estimadores no paramétricos de curvas (Gasser y Müller (1984)). También puede considerarse el *Error cuadrático integrado*:

$$(7) \quad ISE(h) = \int (r_h(x) - r(x))^2 W(x) dx.$$

o el *Error cuadrático medio integrado*, definido como la esperanza de $ISE(h)$. Sin embargo, debido a que r_h es un estimador definido como un cociente, este último error puede no existir, y se debe trabajar entonces con el error

$$(8) \quad MISE^*(h) = E \int \left[\left(\frac{f_h(x)}{1 - F_h(x)} - \frac{f(x)}{1 - F(x)} \right) \frac{1 - F_h(x)}{1 - F(x)} \right]^2 W(x) dx.$$

Estas tres medidas de error son asintóticamente equivalentes (Vieu (1991a)), por lo que es indiferente utilizar una u otra. Este último autor obtiene la siguiente expresión que especifica el comportamiento asintótico del error $MISE^*(h)$:

$$(9) \quad MISE^*(h) = C_1 h^{2k} + C_2 \frac{1}{nh} + o(MISE^*(h)),$$

donde

$$C_1 = \int \left[\left(f^{(k)}(x) + F^{(k)}(x) r(x) \right) \frac{1}{k!} \int u^k K(u) du \right]^2 \frac{W(x) f(x)}{(1 - F(x))^2} dx$$

y

$$C_2 = \int \left(f(x) \int K^2(u) du \right) \frac{W(x) f(x)}{(1 - F(x))^2} dx$$

son constantes reales positivas y $k \geq 2$ es el número de derivadas que se asumen para la función de densidad $f(\cdot)$. El primer término de (9) es debido al sesgo de $r_h(\cdot)$ y el segundo término a su varianza. Podemos derivar dicha expresión y obtener una ventana óptima asintótica, dada por:

$$h_0 = \left(\frac{C_2}{2kC_1} \right)^{1/(2k+1)} n^{-1/(2k+1)}.$$

Debido a que C_1 y C_2 incluyen términos desconocidos, esta ventana es imposible de calcular en la práctica a través de una muestra de datos reales. Una forma de obtener una estimación de tal ventana óptima es utilizar las ideas de validación cruzada de Bowman (1984) y Rudemo (1982) que se basan en calcular h minimizando alguna estimación de

$$ISE(h) = \int r_h^2(x) W(x) dx + \int r^2(x) W(x) dx - 2E \left(\frac{r_h(X)}{1 - F(X)} W(X) \right).$$

En Estévez y Quintela (1999) se propone elegir el h que minimice

$$(10) \quad CV_{l_n}(h) = \int r_h^2(x) W(x) dx - \frac{2}{n} \sum_{i=1}^n \frac{f_h^{-i}(X_i)}{(1 - F_h^{-i}(X_i))(1 - F_n(X_i))} W(X_i),$$

donde $f_h^{-i}(\cdot)$ y $F_h^{-i}(\cdot)$ se construyen como

$$(11) \quad \begin{aligned} f_h^{-i}(x) &= n_{l_n}^{-1} \sum_{|j-i| > l_n} K_h(x, X_j), \\ F_h^{-i}(x) &= n_{l_n}^{-1} \sum_{|j-i| > l_n} H_h(x, X_j), \end{aligned}$$

con l_n un entero positivo, n_{l_n} verificando $nm_{l_n} = \#\{(i, j) / |i - j| > l_n\}$, y las funciones $K_h(\cdot)$ y $H_h(\cdot)$ dadas por $K_h(x, y) = \frac{1}{h} K\left(\frac{x-y}{h}\right)$ y $H_h(x, y) = H\left(\frac{x-y}{h}\right) \cdot f_h^{-i}(x)$ y $F_h^{-i}(x)$ son, de este modo, los estimadores de $f(x)$ y $F(x)$, respectivamente, pero sin utilizar los datos próximos a X_i en el tiempo: $X_{i-l_n}, X_{i-l_n+1}, \dots, X_{i-1}, X_{i+1}, \dots, X_{i+l_n}$. Este tipo de validación cruzada se conoce como *validación cruzada modificada*, y ha sido utilizada en otros contextos de estimación no paramétrica, como la densidad (Hart y Vieu (1990)) o la regresión (Härdle y Vieu (1992), Chu y Marron (1991), Quintela (1994)).

Estévez y Quintela (1999) demuestran que, bajo condiciones habituales en estimación no paramétrica con datos α -mixing, la ventana $\hat{h}(l_n)$ que minimiza $CV_{l_n}(h)$ es asintóticamente óptima, en el sentido de que

$$(12) \quad \frac{d_I(\hat{h}(l_n))}{\inf_h d_I(h)} \rightarrow 1 \text{ casi seguro.}$$

donde $d_I(h) = ISE(h)$, $ASE(h)$ o $MISE^*(h)$.

Sarda y Vieu (1989) demuestran una propiedad similar a (12), considerando el estimador $r_{h,1}(\cdot)$ en vez de $r_h(\cdot)$. En este caso, la función de validación cruzada a minimizar es prácticamente similar a la de (10), y toma la forma

$$(13) \quad CV_{l_n,1}(h) = \int r_{h,1}^2(x) W(x) dx - \frac{2}{n} \sum_{i=1}^n \frac{f_h^{-i}(X_i)}{(1 - F_n(X_i))^2} W(X_i).$$

3. SIMULACIONES

En esta sección realizamos un análisis comparativo, por medio de un estudio de simulación, del comportamiento de los dos estimadores de la función de razón de fallo tratados hasta el momento, $r_h(\cdot)$ y $r_{h,1}(\cdot)$. El estudio realizado es el siguiente:

Hemos generado M muestras, de N observaciones cada una, de un modelo dependiente con distribución Normal y otro con distribución Gamma, del siguiente modo:

Distribución normal

Cada muestra $\{X_t\}_t$ está formada por observaciones de un modelo $AR(1)$ con distribución $N(0, 1)$ obtenidas de la forma:

$$(14) \quad X_t = \rho X_{t-1} + (1 - \rho^2)^{1/2} e_t,$$

donde $\{e_t\}_t$ son variables aleatorias independientes e idénticamente distribuidas según una $N(0, 1)$. El valor de la autocorrelación ρ se ha variado en el conjunto $\{0, \pm 0.3, \pm 0.8\}$, con el fin de ver cómo afecta la dependencia de los datos al procedimiento de validación cruzada.

Distribución gamma

Las observaciones X_t siguen una ley Gamma con parámetros de escala y forma iguales a 1 y 1.5, respectivamente. Una variable aleatoria sigue una distribución Gamma con parámetro de escala β y parámetro de forma α si su función de densidad es:

$$(15) \quad f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \text{ para } x > 0.$$

Las observaciones se obtienen de la forma

$$(16) \quad X_t = \frac{1}{2} \sum_{j=t}^{t+2} Y_j^2,$$

donde Y_1, Y_2, \dots , son variables aleatorias independientes e idénticamente distribuidas según una $N(0, 1)$. De esta manera se obtiene una muestra de datos X_1, \dots, X_n que sigue un modelo 3-dependiente (es decir, los grupos de variables $(X_i, X_{i+1}, \dots, X_j)$ y (X_{j+3}, \dots, X_k) son independientes). Esta condición de dependencia es, obviamente, más restrictiva que la condición fuertemente mixing.

Para cada una de estas distribuciones se ha considerado como función de pesos $W(x) = I_{[-2,2]}(x)$ y $W(x) = I_{[0.5,4.5]}(x)$, respectivamente, y se ha calculado, para cada muestra:

- La ventana que minimiza el error cuadrático promedio (ASE) de $r_h(\cdot)$, que denotaremos por h_{ASE} .
- La ventana que minimiza el error cuadrático promedio (ASE) de $r_{h,1}(\cdot)$, que denotaremos por $h_{ASE,1}$.
- La ventana de validación cruzada, $\hat{h}(l_n)$ para el estimador núcleo $r_h(\cdot)$.
- La ventana de validación cruzada, $\hat{h}_1(l_n)$ para el estimador $r_{h,1}(\cdot)$.

En los dos últimos casos, el valor del parámetro l_n se ha variado desde 0 hasta 10. Para cada una de estas 4 ventanas se ha calculado el error cuadrático ASE asociado. En todos los casos se ha usado como función kernel el núcleo de Epanechnikov:

$$K(u) = \frac{3}{4} (1 - u^2) I_{[-1,1]}(u).$$

Finalmente, hemos calculado las medias y desviaciones típicas de los resultados (amplitudes de banda y errores ASE) sobre las M muestras. Esta última medida aparecerá entre paréntesis.

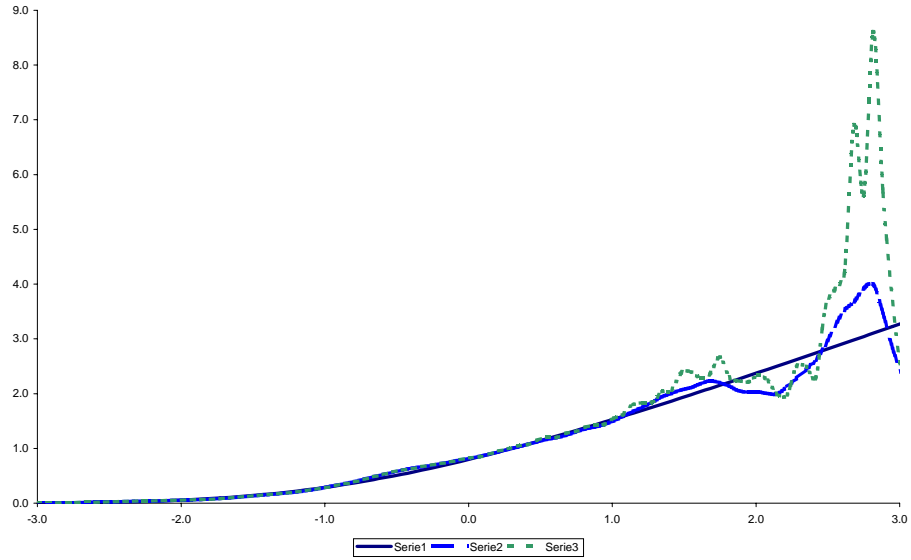


Figura 1. Razón de fallo teórica y estimada. Serie 1: curva teórica. Serie 2: estimación núcleo en el denominador. Serie 3: distribución empírica en el denominador.

En primer lugar, y para hacernos una idea del comportamiento de los estimadores $r_{h,1}(\cdot)$ y $r_h(\cdot)$, presentamos una gráfica, correspondiente a una única muestra de tamaño 1000 de una distribución $N(0, 1)$, generada según (14) con el valor de $\rho = 0.8$. En la figura 1 comparamos los estimadores antes citados con el gráfico de la función de razón de fallo teórica de una distribución normal estándar. Las ventanas, elegidas según los métodos de validación cruzada (10) y (13) son, respectivamente, 0.45 y 0.43 (en ambos casos hemos tomado $l_n = 0$).

En la figura 1 puede apreciarse que, en los valores más altos de la variable, el estimador que utiliza en su denominador la función de distribución empírica se aproxima peor a la función teórica. Esta peor aproximación se aprecia con claridad en las tablas siguientes, donde se ha elegido $M = 200$ y $N = 1000$. Para cada uno de los estimadores y distribuciones, se han calculado las ventanas de validación cruzada y ASE junto con los errores cuadráticos promedio para las distintas estimaciones. Los resultados obtenidos para los valores de $\rho = 0, 0.3$ y 0.8 se muestran en las siguientes tablas:

Distribución normal

Tabla I. $\rho = 0$.

$h_{ASE} = 0.5107$	(0.1323)	$ASE(h_{ASE}) = 0.0060$	(0.0043)
$h_{ASE,1} = 0.5721$	(0.1827)	$ASE(h_{ASE,1}) = 0.0104$	(0.0124)

l_n	$\hat{h}(l_n)$	$ASE(\hat{h}(l_n))$	$\hat{h}_1(l_n)$	$ASE(\hat{h}_1(l_n))$
0	0.4158 (0.1997)	0.0208 (0.0349)	0.6676 (0.3605)	0.0239 (0.0362)
1	0.4147 (0.1985)	0.0206 (0.0343)	0.6569 (0.3616)	0.0247 (0.0383)
2	0.4166 (0.1989)	0.0207 (0.0346)	0.6535 (0.3618)	0.0247 (0.0383)
3	0.4052 (0.1999)	0.0214 (0.0349)	0.6392 (0.3638)	0.0246 (0.0380)
4	0.4059 (0.2018)	0.0216 (0.0352)	0.6466 (0.3649)	0.0251 (0.0386)
5	0.3986 (0.1968)	0.0216 (0.0354)	0.6499 (0.3646)	0.0252 (0.0386)
6	0.3990 (0.1989)	0.0221 (0.0357)	0.6478 (0.3640)	0.0252 (0.0385)
7	0.3945 (0.1980)	0.0224 (0.0359)	0.6415 (0.3648)	0.0251 (0.0382)
8	0.3912 (0.1941)	0.0223 (0.0359)	0.6395 (0.3642)	0.0247 (0.0362)
9	0.3915 (0.1951)	0.0218 (0.0353)	0.6399 (0.3687)	0.0255 (0.0382)
10	0.4007 (0.1927)	0.0207 (0.0342)	0.6652 (0.3566)	0.0226 (0.0330)

Tabla II. $\rho = 0.3$.

$h_{ASE} = 0.4940$	(0.1607)	$ASE(h_{ASE}) = 0.0056$	(0.0046)
$h_{ASE,1} = 0.5355$	(0.1386)	$ASE(h_{ASE,1}) = 0.0107$	(0.0101)

l_n	$\hat{h}(l_n)$	$ASE(\hat{h}(l_n))$	$\hat{h}_1(l_n)$	$ASE(\hat{h}_1(l_n))$
0	0.4972 (0.1967)	0.0174 (0.0244)	0.6524 (0.3340)	0.0244 (0.0290)
1	0.4952 (0.1987)	0.0175 (0.0244)	0.6592 (0.3283)	0.0238 (0.0283)
2	0.4937 (0.1967)	0.0175 (0.0245)	0.6550 (0.3297)	0.0239 (0.0283)
3	0.4905 (0.1978)	0.0177 (0.0249)	0.6608 (0.3378)	0.0247 (0.0293)
4	0.4937 (0.1992)	0.0176 (0.0246)	0.6646 (0.3374)	0.0246 (0.0294)
5	0.4930 (0.2015)	0.0176 (0.0246)	0.6593 (0.3399)	0.0250 (0.0299)
6	0.4902 (0.2031)	0.0177 (0.0246)	0.6550 (0.3410)	0.0247 (0.0294)
7	0.4963 (0.2027)	0.0177 (0.0246)	0.6547 (0.3408)	0.0244 (0.0293)
8	0.4950 (0.2001)	0.0175 (0.0246)	0.6565 (0.3375)	0.0243 (0.0293)
9	0.4993 (0.2001)	0.0174 (0.0245)	0.6566 (0.3367)	0.0236 (0.0282)
10	0.4921 (0.2003)	0.0175 (0.0245)	0.6578 (0.3397)	0.0238 (0.0282)

Tabla III. $\rho = 0.8$.

$h_{ASE} = 0.5894$		(0.2673)	$ASE(h_{ASE}) = 0.0100$		(0.0102)
$h_{ASE,1} = 0.5497$		(0.2345)	$ASE(h_{ASE,1}) = 0.0250$		(0.0271)
<hr/>					
l_n	$\hat{h}(l_n)$	$ASE(\hat{h}(l_n))$	$\hat{h}_1(l_n)$	$ASE(\hat{h}_1(l_n))$	
0	0.4657 (0.1790)	0.0255 (0.0241)	0.6423 (0.3464)	0.0379 (0.0382)	
1	0.4666 (0.1975)	0.0273 (0.0260)	0.6672 (0.3711)	0.0396 (0.0421)	
2	0.4688 (0.1910)	0.0265 (0.0252)	0.6734 (0.3924)	0.0406 (0.0428)	
3	0.4629 (0.1947)	0.0269 (0.0251)	0.6830 (0.4016)	0.0406 (0.0427)	
4	0.4554 (0.1967)	0.0279 (0.0284)	0.6826 (0.4071)	0.0410 (0.0427)	
5	0.4531 (0.1907)	0.0278 (0.0284)	0.6853 (0.4115)	0.0412 (0.0426)	
6	0.4455 (0.1953)	0.0284 (0.0286)	0.6881 (0.4125)	0.0413 (0.0432)	
7	0.4510 (0.1986)	0.0287 (0.0283)	0.6871 (0.4182)	0.0419 (0.0442)	
8	0.4516 (0.1923)	0.0281 (0.0281)	0.6892 (0.4155)	0.0418 (0.0442)	
9	0.4497 (0.1973)	0.0284 (0.0283)	0.6875 (0.4177)	0.0419 (0.0442)	
10	0.4438 (0.1932)	0.0277 (0.0269)	0.6948 (0.4081)	0.0411 (0.0436)	

Los valores de $\rho = -0.3$ y $\rho = -0.8$ en (14) producen resultados similares. Pasamos ahora a exponer los resultados de la distribución Gamma (16):

Tabla IV.

$h_{ASE} = 0.21101$		(0.2190)	$ASE(h_{ASE}) = 0.0018$	(0.0013)
$h_{ASE,1} = 1.3670$		(0.2048)	$ASE(h_{ASE,1}) = 0.0263$	(0.0153)
<hr/>				
l_n	$\hat{h}(l_n)$	$ASE(\hat{h}(l_n))$	$\hat{h}_1(l_n)$	$ASE(\hat{h}_1(l_n))$
0	0.6925 (0.3605)	0.0309 (0.0220)	0.6255 (0.3053)	0.0384 (0.0210)
1	0.8108 (0.3916)	0.0273 (0.0219)	0.8022 (0.3196)	0.0337 (0.0185)
2	0.8373 (0.3835)	0.0260 (0.0214)	0.8115 (0.3225)	0.0336 (0.0191)
3	0.8375 (0.3786)	0.0259 (0.0212)	0.8212 (0.3122)	0.0333 (0.0186)
4	0.8596 (0.3683)	0.0248 (0.0205)	0.8247 (0.3151)	0.0331 (0.0181)
5	0.8560 (0.3724)	0.0250 (0.0207)	0.8244 (0.3178)	0.0331 (0.0181)
6	0.8495 (0.3771)	0.0253 (0.0209)	0.8234 (0.3158)	0.0331 (0.0181)
7	0.8504 (0.3768)	0.0253 (0.0209)	0.8135 (0.3214)	0.0335 (0.0185)
8	0.8308 (0.3716)	0.0258 (0.0210)	0.8115 (0.3206)	0.0335 (0.0185)
9	0.8415 (0.3725)	0.0255 (0.0209)	0.8131 (0.3228)	0.0335 (0.0185)
10	0.8457 (0.3709)	0.0253 (0.0208)	0.8095 (0.3246)	0.0335 (0.0184)

Desde el punto de vista comparativo de los dos estimadores, observamos que el método de validación cruzada tiende a producir, en el caso de la distribución Normal, una infra-suavización de la ventana óptima h_{ASE} para $r_h(\cdot)$ y, sin embargo, una sobresuavización de $h_{ASE,1}$ para $r_{h,1}(\cdot)$. Para la distribución Gamma, no obstante, dicho método infra-suaviza las ventanas para ambos estimadores. Observamos también que, para cada valor de ρ y cada l_n , los errores $ASE(\hat{h}(l_n))$ son menores, en media, que los correspondientes $ASE(\hat{h}_1(l_n))$ y con dispersiones más pequeñas. En resumen, parece que el uso del estimador núcleo va a proporcionar buenas estimaciones, no sólo mejores que las de $r_{h,1}(\cdot)$, sino también en comparación con la estimación óptima, $r_{h_{ASE}}(\cdot)$.

Si observamos los resultados para cada estimador por separado, nos damos cuenta de que el valor del parámetro l_n en el procedimiento de validación cruzada tiene un efecto muy débil, al contrario de lo que ocurre en otros problemas de estimación. En contextos similares, como en estimación de densidad (Hart y Vieu (1990)) o regresión (Quintela (1994)), cuando trabajamos, por ejemplo, con datos que presentan autocorrelación positiva, interesa aumentar el valor de l_n a medida que ésta crece. Aquí, bajo el modelo $AR(1)$ con distribución marginal $N(0, 1)$, se observa que un incremento de l_n prácticamente no modifica el valor de la amplitud de banda obtenida para $l_n = 0$. Para $\rho = 0.8$ por ejemplo, las ventanas seleccionadas por validación cruzada se mueven en el intervalo $[0.44, 0.47]$, sin observarse una tendencia clara al aumentar l_n (más detalles sobre esta afirmación pueden consultarse en Estévez y Quintela (1999)). Sin embargo, bajo la distribución Gamma, sí se logran ventanas ligeramente distintas al utilizar $l_n = 0$ o $l_n > 0$, aunque tal diferencia tampoco conlleva grandes mejorías en lo que a errores de estimación se refiere. La explicación de este hecho está en la forma (simétrica a la J) de la función de validación cruzada.

4. SELECCIÓN LOCAL DEL PARÁMETRO VENTANA

En la estimación no paramétrica de curvas es un fenómeno relativamente frecuente que la utilización de un único parámetro ventana, para la estimación de una curva en concreto, no presente buenos resultados. Así, en regiones de baja densidad de datos, es preferible elegir un parámetro ventana grande a fin de basar la estimación en un número suficientemente grande de puntos. Mientras, en aquellas áreas en las que existen un mayor número de datos, es preferible elegir parámetros ventana pequeños, pues evidenciarán más claramente las características locales de la función a estimar. Con esta filosofía se construyen los estimadores de ventana local, que utilizan una ventana diferente h_x dependiendo del punto concreto x donde se realice la estimación. Básicamente, se pretende seleccionar una ventana h_x que minimice algún estimador del *Error cuadrático medio puntual*:

$$(17) \quad MSE_x(h) = E(r_h(x) - r(x))^2.$$

En esta sección definimos un método de selección local de la ventana que se demuestra asintóticamente óptimo en el sentido de (12), con (12) adaptado a un método de validación cruzada de tipo local. Concretamente, para cada punto x de un intervalo $[a, b] \subset \text{Soporte}(X)$ ($\text{Soporte}(X) = \{x \in \mathbb{R} / f(x) > 0\}$), elegiremos una ventana h_x minimizando la función

$$(18) \quad CV_{x,ln}(h) = \int_a^b r_h^2(y) W_{n,x}(y) dy - 2n^{-1} \sum_{i=1}^n \frac{f_h^{-i}(X_i)}{(1 - F_h^{-i}(X_i))(1 - F_n(X_i))} W_{n,x}(X_i),$$

donde $f_h^{-i}(\cdot)$ y $F_h^{-i}(\cdot)$ se definen igual que en (11), y $W_{n,x}(\cdot)$ es una sucesión de funciones de peso centradas en el punto x . Es habitual elegir funciones del tipo

$$(19) \quad W_{n,x}(y) = \begin{cases} 1 & \text{si } y \in [x - t_{0,n}, x + t_{0,n}] \\ 0 & \text{si } y \notin [x - t_{0,n}, x + t_{0,n}] \end{cases}$$

siendo $t_{0,n}$ un número real que tiende a 0 cuando n tiende a infinito, y dependiente de x (Vieu (1991b), Quintela y Vilar (1992)). Como puede verse en estos trabajos, la elección de esta función $W_{n,x}(\cdot)$ no influye de manera especial en el procedimiento de selección de la ventana.

Para el establecimiento de un resultado de optimalidad asintótica para el procedimiento de validación cruzada local utilizaremos las siguientes hipótesis:

(H.1) $\{X_n\}_{n \in \mathbb{N}}$ es una sucesión estrictamente estacionaria de variables α -mixing.

(H.2) La sucesión de pesos locales $W_{n,x}(\cdot)$ verifica:

$$W_{n,x}(u) \geq 0, \quad \int W_{n,x}(u) du = 1,$$

$$\sup_{x,u} W_{n,x}(u) = O\left(n^\beta\right) \quad \text{con } \beta < \frac{1-2ku}{2}.$$

(H.3) Para toda función $g(\cdot)$ continua se verifica:

$$\int W_{n,x}(u) g(u) du \rightarrow g(x) \quad \text{uniformemente para } x \in [a, b].$$

(H.4) Existe $\gamma > 0$ tal que $\forall x \in [a, b], 1 - F(x) > \gamma$ y $f(x) \geq \gamma$.

(H.5) Existe $M_1, 0 < M_1 < \infty$, tal que $f(u) \leq M_1 \forall u \in [a, b]$.

(H.6) La densidad $f(\cdot)$ tiene $k(\geq 2)$ derivadas continuas.

(H.7) Para todo $j \geq 2$ existe la función de densidad conjunta $f_{1,j}(\cdot, \cdot)$ y satisface:

$$\sup_{u,v \in [a,b]} \int |f_{1,j}(u,v) - f(u)f(v)| du dv \leq \alpha(j-1).$$

(H.8) La función núcleo $K(\cdot)$ es una función de soporte compacto que encierra área 1, es decir, $\int K(u) du = 1$.

(H.9) $K(\cdot)$ es de orden k , esto es

$$\int u^m K(u) du = 0, \text{ para } m = 1, \dots, k-1 \text{ y } 0 < \int u^k K(u) du < \infty.$$

(H.10) $K(\cdot)$ es Lipschitz continua, es decir, $\exists C_K \ 0 < C_K < \infty$ tal que

$$|K(x) - K(y)| \leq C_K |x - y| \quad \forall x, y \in \mathbb{R}.$$

(H.11) La transformada de Fourier de $K(\cdot)$ es absolutamente integrable.

(H.12) $h \in H_n = [A_1 n^{-u}, A_2 n^{-v}]$, $0 < A_1 < A_2 < \infty$, $0 < v \leq 1/(2k+1) \leq u < 2/(4k+1)$.

(H.13) $0 \leq l_n \leq C n^{r_1}$, con $0 < r_1 < 1 - \beta - u(4k+1)/2$.

(H.14) Para

$$r_2 = U + V + (2\beta + 2u + 4ku)(2 + \frac{U}{V}),$$

$$U = 1 + 2u + 2ku + \beta - v, \quad V = 2 - 2\beta - u(1 + 4k) - 2r_1,$$

se tiene que

$$\tilde{\alpha}(n^{r_1}) = \sup_{t > n^{r_1}} \alpha(t) = o(n^{-r_2}).$$

Teorema 4.1. *Bajo las hipótesis (H.1)-(H.14), se tiene que*

$$\sup_{x \in [a,b]} \frac{d_x(\hat{h}_x(l_n))}{\inf_{h \in H_n} d_x(h)} \rightarrow 1 \quad \text{casi seguro},$$

donde $\hat{h}_x(l_n)$ es la ventana de validación cruzada local y $d_x(h)$ es $MSE_x(h)$ (17) o cualquiera de las versiones locales de los errores cuadráticos (6), (7) y (8):

$$ASE_x(h) = \frac{1}{n} \sum_{i=1}^n (r_h(X_i) - r(X_i))^2 W_{n,x}(X_i),$$

$$ISE_x(h) = \int_a^b (r_h(y) - r(y))^2 W_{n,x}(y) dy$$

y

$$MISE_x^*(h) = E \int \left[\left(\frac{f_h(y)}{1 - F_h(y)} - \frac{f(y)}{1 - F(y)} \right) \frac{1 - F_h(y)}{1 - F(y)} \right]^2 W_{n,x}(y) dy.$$

Comentario 4.1. Las hipótesis bajo las que se establece este teorema son del mismo tipo de las utilizadas en los trabajos de Hart y Vieu (1990) y Quintela y Vilar (1992), por lo que sus comentarios sobre las mismas siguen siendo válidos en nuestro contexto. Estas hipótesis son, a su vez, las mismas que las utilizadas en el teorema 6 de Estévez y Quintela (1999), a las que se han añadido las correspondientes a la utilización de las funciones de peso locales $W_{n,x}(\cdot)$.

La demostración de este teorema sigue los mismos pasos que el teorema 6 de Estévez y Quintela (1999), pero teniendo en cuenta las hipótesis **(H.2)** y **(H.3)** sobre los pesos locales. La demostración hace uso, también, de los siguientes resultados sobre equivalencia asintótica de los errores cuadráticos locales y de la descomposición asintótica de los mismos:

Teorema 4.2. Bajo las hipótesis del Teorema 4.1, se tiene que

$$\sup_{h \in H_n} \frac{|d_x(h) - d'_x(h)|}{d'_x(h)} \rightarrow 0 \quad \text{casi seguro,}$$

y

$$d_x(h) = C_1 h^{2k} + C_2 \frac{1}{nh} + o\left(h^{2k} + \frac{1}{nh}\right),$$

para $d_x(h)$ y $d'_x(h)$ cualesquiera de las funciones $MSE_x(h)$, $ASE_x(h)$, $ISE_x(h)$ o $MISE_x^*(h)$.

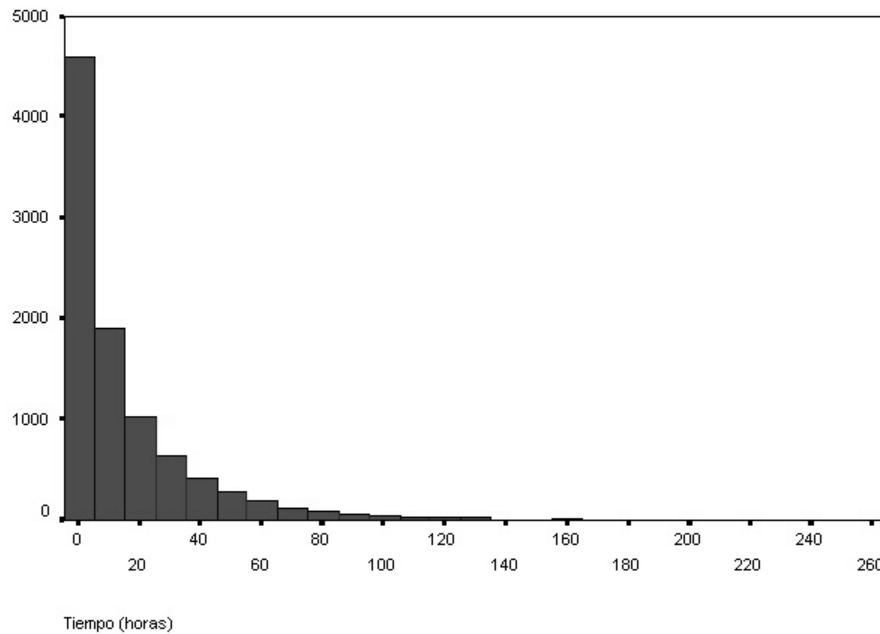


Figura 2. Histograma de la serie de datos de California.

La demostración de los teoremas 4.1 y 4.2 se ha resumido en el Apéndice del presente trabajo.

5. APLICACIONES A DATOS SÍSMICOS

El modelo estadístico más simple para ajustar una serie de tiempo de ocurrencia entre terremotos es el proceso de Poisson (Vere-Jones (1970)). Bajo este modelo, los intervalos de tiempo entre eventos consecutivos están distribuidos exponencialmente, de modo que la función de riesgo es una línea recta de altura igual al inverso del valor medio de la variable.

Sin embargo, este modelo presupone independencia entre los eventos, de modo que la ocurrencia de un terremoto no estaría influenciada por la de los terremotos previos, lo que obviamente resulta una hipótesis bastante alejada de la realidad.

La desviación de las observaciones de este modelo se atribuye, principalmente, a la existencia de un número muy grande de intervalos de tiempo pequeños, debido a que los terremotos no ocurren como sucesos aislados, sino que tienden a agruparse en «clusters», es decir, un movimiento concreto viene seguido y/o precedido, habitualmente, por otra serie de «replicas» y/o «precursores» (Lomnitz y Hax (1967), Vere-Jones y Davies (1966)).

Un modelo estadístico de tipo paramétrico que ajusta más convenientemente esta tendencia al agrupamiento o «clustering» de los movimientos sísmicos es la distribución Gamma ((Vere-Jones (1970), Udias y Rice (1975)).

Los datos de nuestro primer ejemplo corresponden a un fichero con la información de los terremotos ocurridos en California (Estados Unidos), desde el 1 de enero de 1980 hasta el 12 de octubre de 1996 (inclusive). De ellos, hemos seleccionado aquellos que tienen magnitud mayor o igual que 3 en la escala de Richter, y profundidad menor o igual que 50 km. De esta forma, estamos considerando aquellos terremotos cuya localización goza de mayor precisión. Considerando el tiempo de ocurrencia de cada uno de estos movimientos sísmicos, construimos la serie de tiempo $\{\Delta t_i\}_{i=1}^n$ de intervalos de tiempo (en horas) entre movimientos consecutivos. Se obtiene una muestra de tamaño $n = 9471$ datos.

En primer lugar mostramos, en la figura 2, un histograma de los datos $\{\Delta t_i\}_{i=1}^n$. En éste puede observarse el efecto de agrupamiento antes citado, al concentrarse una gran cantidad de datos en valores próximos a cero. Un estudio descriptivo de los mismos arroja, como valores significativos: media muestral = 15.53, mediana = 5.48, desviación típica = 24.204, mínimo de la variable = 0.00027 y máximo de la variable = 256.2886.

Un ajuste de tipo paramétrico de la densidad de estos datos a una distribución Gamma como en (15) proporciona unos valores para α y β de 0.4117 y 0.02651, respectivamente. Sin embargo, tanto el test Chi-cuadrado como el de Kolmogorov-Smirnov de bondad de ajuste arrojan valores para el p-valor del test iguales o muy próximos a cero, indicando que debe rechazarse la hipótesis de pertenencia de los datos a la distribución especificada.

Para realizar una estimación no paramétrica de la función de riesgo asociada a la muestra $\{\Delta t_i\}_{i=1}^n$, aplicamos el algoritmo de validación cruzada (10) a dichos datos. Debido a la escasa influencia del valor l_n , lo hemos tomado igual a cero. El valor de la ventana así obtenida es $\hat{h}(0) = 0.25$ (horas), y la gráfica de la función de razón de fallo $r_h(\cdot)$ estimada con dicho parámetro aparece en la figura 3. En la misma, hemos incluido la gráfica de la función de razón de fallo que correspondería a una distribución Gamma de parámetros los estimados a partir de la muestra ($\alpha = 0.4117$ y $\beta = 0.02651$).

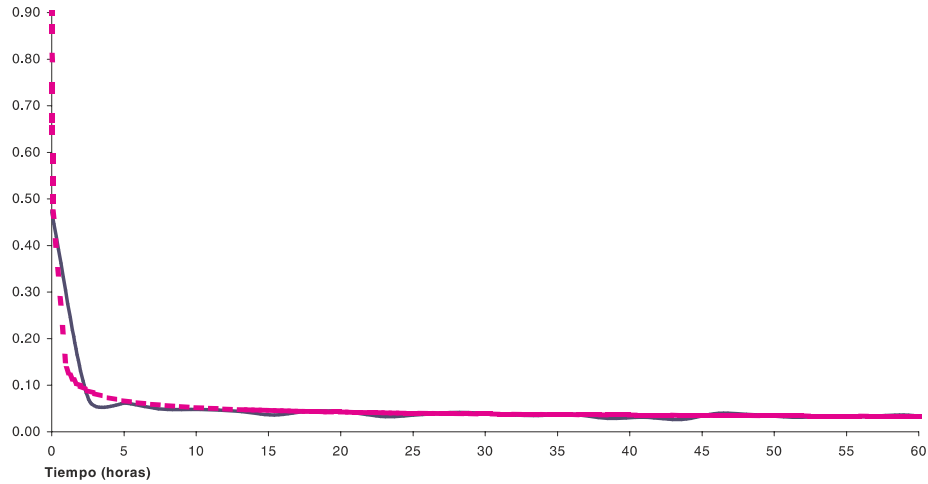


Figura 3. Estimaciones paramétrica y no paramétrica con los datos de California. Línea sólida: estimación no paramétrica. Línea punteada: estimación paramétrica.

Atendiendo a la estimación no paramétrica, podemos observar el elevado riesgo que existe en los primeros momentos, y como la función tiende a ser constante a partir de las 6 horas, aproximadamente. En la gráfica, el eje de abscisas sólo alcanza hasta el valor 60 para una mejor visualización, puesto que a partir de dicho valor sólo existen un cinco por ciento del total de los datos. De hecho, el cuantil de orden 90 ya es el valor 44.29 horas. Un efecto decreciente como el de esta figura puede observarse en las figuras 8 y 9 de Rice y Rosenblatt (1976), donde se analiza un conjunto de terremotos ocurridos en la Falla de San Andreas, una región comprendida en nuestro espacio geográfico de análisis.

La comparación con la curva ajustada a una distribución Gamma revela la potente capacidad de la estimación no paramétrica para una estimación más correcta del riesgo instantáneo de ocurrencia de un terremoto. En efecto, la curva que correspondería a una distribución Gamma indicaría unos valores del riesgo no aceptables en los valores del tiempo más próximos a cero (el eje de ordenadas es una asíntota de la curva ajustada paramétricamente), mientras que ambas estimaciones son muy similares a partir de las 5 o 6 horas, aproximadamente.

El segundo ejemplo que hemos considerado corresponde a la serie de tiempo formada por los intervalos de tiempo (en horas) entre la ocurrencia de terremotos consecutivos en Granada, durante el periodo de tiempo comprendido desde enero de 1993 hasta diciembre de 1995 (datos obtenidos del Instituto Geográfico Nacional). Esta serie de tiempo $\{\Delta t_i\}_{i=1}^n$ se compone de $n = 172$ datos.

Aplicando el procedimiento de validación cruzada global a esta serie de tiempo, obtenemos como valores para las ventanas $\hat{h}(0) = 38.2$, $\hat{h}(1) = 38.0$, $\hat{h}(2) = 31.0$, $\hat{h}(3) = 30.2$, $\hat{h}(4) = 30.4$ y $\hat{h}(5) = 29.8$ que proporcionan, prácticamente, las mismas estimaciones. En la figura 5 (comentada más adelante) se muestra una gráfica de la curva estimada con el parámetro $\hat{h}(0)$.

Para comprobar los resultados prácticos del procedimiento de selección de la ventana mediante validación cruzada local, aplicamos el método definido en (18) a este conjunto de datos. Para ello, hemos realizado una partición del intervalo muestral en 33 puntos equiespaciados y, para cada uno de ellos, hemos calculado una ventana local. Como funciones de peso hemos tomado (19), con $t_{0,n}$ la mitad de la desviación típica de los datos. El valor de l_n se eligió igual a cero. Una representación gráfica de las diferentes ventanas locales obtenidas $\hat{h}_x(0)$ se presenta en la figura 4.

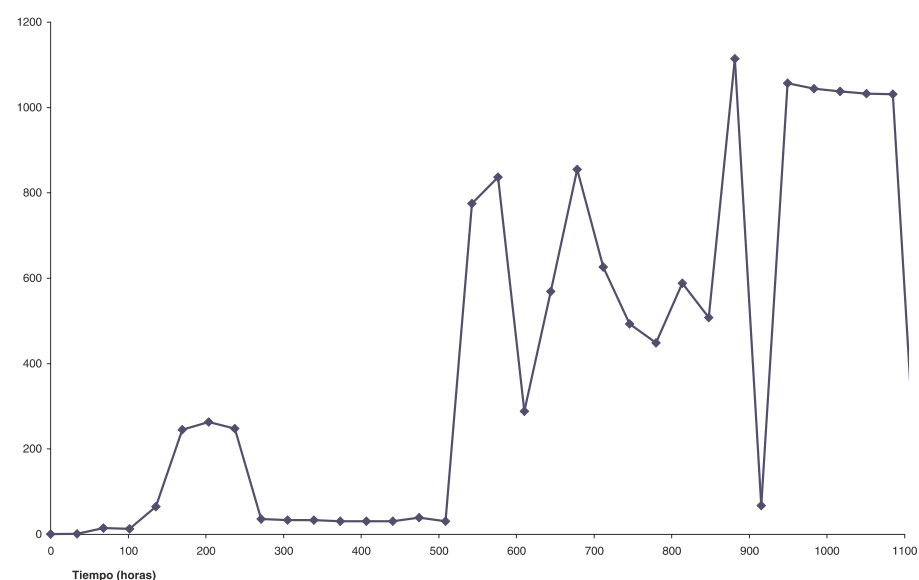


Figura 4. Ventanas locales.

En la misma podemos observar la tendencia a obtenerse ventanas más grandes a medida que el tiempo crece, puesto que la dispersión de los datos es también mayor. La figura 5 muestra las estimaciones locales y globales de la función de riesgo para los datos en estudio:

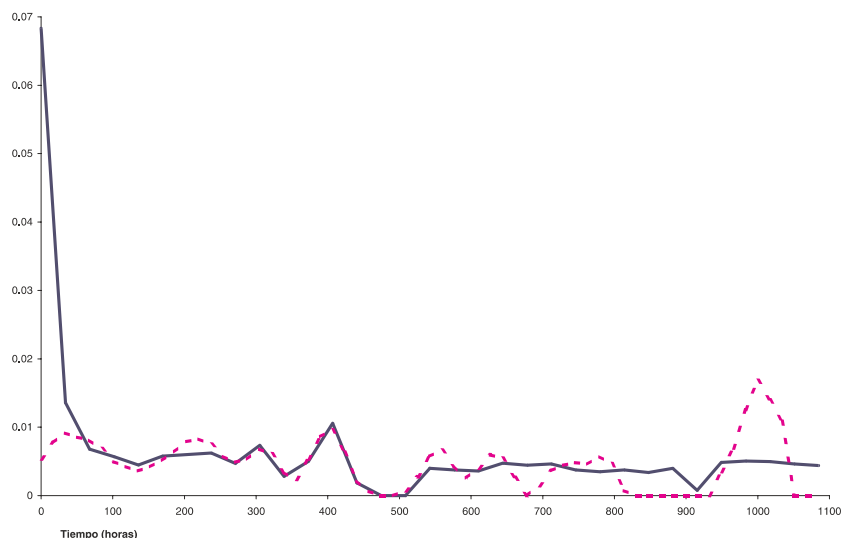


Figura 5. Estimaciones global y local de la razón de fallo con los datos de Granda. Línea sólida: estimación local. Línea punteada: Estimación global.

Puede observarse que los estimadores son similares lejos de las fronteras. Para intervalos de tiempo grandes, el estimador local corrige los valores excesivos que muestra la estimación global. Igualmente, en los intervalos próximos a cero también se corrige la mala estimación causada por una ventana global que resulta excesivamente grande en dicho extremo. Este hecho pone también de manifiesto el enorme interés que puede tener una estimación del riesgo de tipo local cuando se maneja una muestra con no demasiados datos. En el primer caso analizado, el de los terremotos de California, la estimación global produjo resultados muy aceptables gracias a que la muestra contenía un número de datos muy grande. En el segundo ejemplo, el tamaño de muestra pequeño produjo una estimación global del riesgo muy alejada de la realidad en los valores próximos a cero, y este problema pudo evitarse mediante una estimación de tipo local.

6. APÉNDICE: DEMOSTRACIONES DE LOS RESULTADOS RELATIVOS AL USO DE VENTANAS LOCALES

En la presente sección se dan las demostraciones resumidas de los resultados de la sección 4, relativos al uso de ventanas locales en estimación núcleo de la función de razón de fallo. Las pruebas detalladas pueden ser vistas en Estévez (2001).

Demostración del Teorema 4.1. La demostración es muy similar a la correspondiente del caso global (Estévez y Quintela (1999) prueba de su Teorema 6), por lo que nos ceñiremos a una revisión superficial de la misma. La diferencia entre ambas pruebas estriba en la función de pesos; en tanto en el Teorema 6 de Estévez y Quintela (1999) se hacía uso de una función acotada, ahora se dispone de una sucesión de funciones $W_{n,x}(\cdot)$ verificando: $\sup_{x,u} W_{n,x}(u) = O(n^\beta)$ con $\beta < \frac{1-2ku}{2}$, es decir, tienen la posibilidad de tender a infinito con el tamaño muestral aunque de un modo muy lento.

Al igual que en el caso global, la equivalencia asintótica entre los errores locales $ASE_x(h)$, $ISE_x(h)$, $MISE_x^*(h)$ y $MSE_x^*(h)$ (Teorema 4.2) permite demostrar el presente resultado probando, para cada $x \in [a, b]$, la existencia de una v.a. T , independiente de h , que satisfaga:

$$(20) \quad \sup_{h \in H_n} \frac{|ISE_x(h) - CV_{x,l_n}(h) - T|}{ISE_x(h)} \rightarrow 0 \quad c.s..$$

Para ello, se tratará de localizar la v.a. T y, simultáneamente, lograr una descomposición de fácil manejo para $ISE_x(r_h) - CV_{l_n,x}(h)$. En primer lugar, de la definición de las funciones $ISE_x(h)$ y $CV_{x,l_n}(h)$ se obtiene que

$$(21) \quad \begin{aligned} ISE_x(r_h) - CV_{x,l_n}(h) &= \int_a^b \frac{f^2(y)}{(1-F(y))^2} W_{n,x}(y) dy - \\ &- 2 \int_a^b \frac{f_h(y)}{(1-F_h(y))(1-F(y))} W_{n,x}(y) f(y) dy + \\ &+ \frac{2}{n} \sum_{i=1}^n \frac{f_h^{-i}(X_i)}{(1-F_h^{-i}(X_i))(1-F_n(X_i))} W_{n,x}(X_i). \end{aligned}$$

Ahora bien, por analogía con el caso global, utilizando el Teorema 4.2 junto con el hecho de que $f(\cdot)$ está acotada en el intervalo $[a, b]$ y las hipótesis [H.2] y [H.4], se obtiene la siguiente descomposición para el segundo sumando de (21):

$$\begin{aligned} &\int_a^b \frac{f_h(y)}{(1-F_h(y))(1-F(y))} W_{n,x}(y) f(y) dy = \\ &= \int_a^b \frac{f_h(y)}{(1-F(y))^2} W_{n,x}(y) f(y) dy - \int_a^b \frac{f(y)(F(y) - F_h(y))}{(1-F(y))^3} W_{n,x}(y) f(y) dy - \end{aligned}$$

$$\begin{aligned}
(22) \quad & - \int_a^b \frac{(f_h(y) - f(y))(F(y) - EF_h(y))}{(1 - F(y))^3} W_{n,x}(y) f(y) dy - \\
& - \int_a^b \frac{(f_h(y) - f(y))(EF_h(y) - F_h(y))}{(1 - F(y))^3} W_{n,x}(y) f(y) dy + \\
& + \int_a^b \frac{f(y)(F(y) - F_h(y))^2}{(1 - F(y))^4} W_{n,x}(y) f(y) dy + o(ISE_x(h)) \quad c.s..
\end{aligned}$$

De forma análoga, el tercer sumando de (21) se puede escribir como

$$\begin{aligned}
& \frac{2}{n} \sum_{i=1}^n \frac{f_h^{-i}(X_i)}{(1 - F_h^{-i}(X_i))(1 - F_n(X_i))} W_{n,x}(X_i) = \\
& = \frac{2}{n} \sum_{i=1}^n \frac{f_h^{-i}(X_i)}{(1 - F(X_i))^2} W_{n,x}(X_i) + T' - \frac{2}{n} \sum_{i=1}^n \frac{f(X_i)(F(X_i) - F_h^{-i}(X_i))}{(1 - F(X_i))^3} W_{n,x}(X_i) - \\
(23) \quad & - \frac{2}{n} \sum_{i=1}^n \frac{(f_h^{-i}(X_i) - f(X_i))(F(X_i) - \bar{E}[F_h(X_i)/X_i])}{(1 - F(X_i))^3} W_{n,x}(X_i) - \\
& - \frac{2}{n} \sum_{i=1}^n \frac{(f_h^{-i}(X_i) - f(X_i))(\bar{E}[F_h(X_i)/X_i] - F_h^{-i}(X_i))}{(1 - F(X_i))^3} W_{n,x}(X_i) + \\
& + \frac{2}{n} \sum_{i=1}^n \frac{f(X_i)(F(X_i) - F_h^{-i}(X_i))^2}{(1 - F(X_i))^4} W_{n,x}(X_i) + o(ISE_x(h)) \quad c.s.,
\end{aligned}$$

donde

$$T' = \frac{2}{n} \sum_{i=1}^n \frac{f(X_i)(F_n(X_i) - F(X_i))}{(1 - F(X_i))^3} W_{n,x}(X_i)$$

y $\bar{E}[F_h(X_i)/X_i]$ es la media condicional si los datos fuesen independientes. Es importante reseñar que si bien la prueba de (22) se obtiene, en virtud de la condición $\int W_{n,x}(u) du = 1$, trabajando como si hubiese una única función de pesos $W(\cdot)$, la demostración de (23) requiere más cautela y el uso de $\sup_{x,u} W_{n,x}(u) = O(n^\beta)$ con $\beta <$

$\frac{1-2ku}{2}$. Así pues, recopilando (21)-(23), se obtiene la siguiente factorización para el numerador de (20):

$$(24) \quad ISE_x(h) - CV_{x,l_n}(h) - T = 2CT_{l_n,x}^{(1)}(h) - 2CT_{l_n,x}^{(2)}(h) - 2CT_{l_n,x}^{(3)}(h) - \\ - 2CT_{l_n,x}^{(4)}(h) + 2CT_{l_n,x}^{(5)}(h) + o(ISE_x(h)) \quad c.s.,$$

donde la v.a. T está dada por

$$T = \frac{2}{n} \sum_{i=1}^n \frac{f(X_i)}{(1-F(X_i))^2} W_{n,x}(X_i) - \int_a^b \frac{f^2(y)}{(1-F(y))^2} W_{n,x}(y) dy + \\ + \frac{2}{n} \sum_{i=1}^n \frac{f(X_i)(F_n(X_i) - F(X_i))}{(1-F(X_i))^3} W_{n,x}(X_i)$$

y los términos $CT_{l_n,x}^{(i)}(h)$ son:

$$CT_{l_n,x}^{(1)}(h) = \frac{1}{n} \sum_{i=1}^n \frac{f_h^{-i}(X_i) W_{n,x}(X_i)}{(1-F(X_i))^2} - \frac{1}{n} \sum_{i=1}^n \frac{f(X_i) W_{n,x}(X_i)}{(1-F(X_i))^2} + \\ + \int_a^b \frac{f^2(y)}{(1-F(y))^2} W_{n,x}(y) dy - \int_a^b \frac{f_h(y)}{(1-F(y))^2} f(y) W_{n,x}(y) dy;$$

$$CT_{l_n,x}^{(2)}(h) = \frac{1}{n} \sum_{i=1}^n \frac{(f_h^{-i}(X_i) - f(X_i))(F(X_i) - \bar{E}[F_h(X_i)/X_i])}{(1-F(X_i))^3} W_{n,x}(X_i) - \\ - \int_a^b \frac{(f_h(y) - f(y))(F(y) - EF_h(y))}{(1-F(y))^3} W_{n,x}(y) f(y) dy;$$

$$CT_{l_n,x}^{(3)}(h) = \frac{1}{n} \sum_{i=1}^n \frac{(f_h^{-i}(X_i) - f(X_i))(\bar{E}[F_h(X_i)/X_i] - F_h^{-i}(X_i))}{(1-F(X_i))^3} W_{n,x}(X_i) - \\ - \int_a^b \frac{(f_h(y) - f(y))(EF_h(y) - F_h(y))}{(1-F(y))^3} W_{n,x}(y) f(y) dy;$$

$$CT_{l_n, x}^{(4)}(h) = \frac{1}{n} \sum_{i=1}^n \frac{f(X_i) (F(X_i) - F_h^{-i}(X_i))}{(1 - F(X_i))^3} W_{n, x}(X_i) -$$

$$- \int_a^b \frac{f(y) (F(y) - F_h(y))}{(1 - F(y))^3} W_{n, x}(y) f(y) dy$$

y

$$CT_{l_n, x}^{(5)}(h) = \frac{1}{n} \sum_{i=1}^n \frac{f(X_i) (F(X_i) - F_h^{-i}(X_i))^2}{(1 - F(X_i))^4} W_{n, x}(X_i) -$$

$$- \int_a^b \frac{f(y) (F(y) - F_h(y))^2}{(1 - F(y))^4} W_{n, x}(y) f(y) dy.$$

Por consiguiente, la demostración de (20) es consecuencia directa de (24) y

$$(25) \quad \sup_{h \in H_n} \frac{|CT_{l_n, x}^{(i)}(h)|}{ISE_x(h)} \rightarrow 0 \quad c.s., \text{ para } i = 1, \dots, 5.$$

Para probar (25) seguiremos, como hasta el momento, esencialmente los mismos pasos que en el procedimiento de validación cruzada global, notando que también, en virtud del Teorema 4.2, $ISE_x(h) \geq Ch^{2k} \left(C \frac{1}{nh} \right) c.s..$

Prueba de (25) para $i = 1$ ($i = 2$). Comparando $CT_{l_n, x}^{(1)}(h)$ (respectivamente $CT_{l_n, x}^{(2)}(h)$) con el término $T_x(h)$ de la demostración del Teorema 3.1.2 de Quintela (1992), se observa que son idénticos salvo un pequeño matiz: la sucesión de funciones de ponderación. Mientras el autor utiliza una sucesión de funciones $W_{n, x}(\cdot)$ que cumple [H.2] y [H.3], en nuestro caso aparece

$$\frac{W_{n, x}(\cdot)}{(1 - F(\cdot))^2} \left(\text{respectivamente } \frac{B(\cdot) W_{n, x}(\cdot)}{(1 - F(\cdot))^3}, \text{ con } B(y) = \right.$$

$$\left. = F(y) - \int H\left(\frac{y-u}{h}\right) f(u) du \right)$$

que, por otra parte, también satisface las condiciones anteriores. Así pues, notando que los órdenes de convergencia de los errores $MISE_x(\cdot)$ para densidad y razón de fallo coinciden (ver Teoremas 4.2.4 y 4.2.6 en Estévez (2001)), su prueba se adapta de forma trivial a nuestro contexto.

Prueba de (25) para $i = 5$. Así como en el caso global, el término $CT_{l_n, x}^{(5)}(h)$ puede ser acotado tal y como sigue:

$$\begin{aligned} \left| CT_{l_n, x}^{(5)}(h) \right| &\leq \left| \frac{1}{n} \sum_{i=1}^n \frac{f(X_i) (F(X_i) - F_h(X_i))^2}{(1 - F(X_i))^4} W_{n, x}(X_i) - \right. \\ &\quad \left. - \int_a^b \frac{f(y) (F(y) - F_h(y))^2}{(1 - F(y))^4} W_{n, x}(y) f(y) dy \right| + \\ &\quad + \left| \frac{1}{n} \sum_{i=1}^n \frac{f(X_i) (F_h(X_i) - F_h^{-i}(X_i))^2}{(1 - F(X_i))^4} W_{n, x}(X_i) \right| + \\ &\quad + \left| \frac{2}{n} \sum_{i=1}^n (F(X_i) - F_h(X_i)) (F_h(X_i) - F_h^{-i}(X_i)) \frac{f(X_i)}{(1 - F(X_i))^4} W_{n, x}(X_i) \right|, \end{aligned}$$

de lo que se deriva (25) sin más que utilizar la equivalencia asintótica entre los errores locales $ASE_x(F_h)$ e $ISE_x(F_h)$ (ver Teorema 4.2.5 en Estévez (2001)), el Lema 2 de Estévez y Quintela (1999), la condición **[H.13]** y el hecho de que $F_h(y) - F_h^{-i}(y) = O(n^{r_1-1})$ c.s..

Prueba de (25) para $i = 3$. Con un modo de proceder análogo al del término precedente, se obtiene que

$$\begin{aligned} \left| CT_{l_n, x}^{(3)}(h) \right| &\leq \left| \frac{1}{n} \sum_{i=1}^n \frac{(f_h^{-i}(X_i) - f(X_i)) (\bar{E}[F_h(X_i) / X_i] - F_h(X_i))}{(1 - F(X_i))^3} W_{n, x}(X_i) \right| + \\ &\quad + \left| \frac{1}{n} \sum_{i=1}^n \frac{(f_h^{-i}(X_i) - f(X_i)) (F_h(X_i) - F_h^{-i}(X_i))}{(1 - F(X_i))^3} W_{n, x}(X_i) \right| + \\ &\quad + \left| \int_a^b \frac{(f_h(y) - f(y)) (EF_h(y) - F_h(y))}{(1 - F(y))^3} W_{n, x}(y) f(y) dy \right|, \end{aligned}$$

que da lugar, trabajando como en el caso global, a (25) sin más que aplicar los Lemmas 2, 3 y 12 de Estévez y Quintela (1999) y las hipótesis **[H.12]** y **[H.13]**.

Prueba de (25) para $i = 4$. Llegado a este punto de la demostración, interesa considerar H'_n un subconjunto finito de H_n formado por puntos equiespaciados y tal que $\#H'_n = n^\rho$, siendo ρ un número real verificando $2u(1+k) + \beta - v < \rho < 2u(1+k) + \beta - v + V$.

Además, para cada $h \in H_n$ denotaremos por h^* el punto de H'_n más próximo a h . Pues bien, con esta notación la prueba de (25) para $i = 4$ es consecuencia inmediata de los siguientes lemas:

Lema 6.1. *Bajo las hipótesis del Teorema 4.1, se verifica que:*

$$\sup_{h \in H'_n} \frac{|CT_{l_n^*, x}^{(4)}(h)|}{ISE_x(h)} \rightarrow 0 \quad c.s..$$

Lema 6.2. *Bajo las hipótesis del Teorema 4.1, se verifica que:*

$$\sup_{h \in H_n} \frac{|CT_{l_n^*, x}^{(4)}(h) - CT_{l_n^*, x}^{(4)}(h^*)|}{ISE_x(h)} \rightarrow 0 \quad c.s..$$

Lema 6.3. *Bajo las hipótesis del Teorema 4.1, se verifica que:*

$$\sup_{h \in H_n} \frac{|CT_{l_n, x}^{(4)}(h) - CT_{l_n^*, x}^{(4)}(h)|}{ISE_x(h)} \rightarrow 0 \quad c.s..$$

Demostración del Lema 6.3. No es más que una réplica de la prueba del Lema 14 de Estévez y Quintela (1999) con la salvedad de que en lugar de su condición **[H.13]** se utiliza $\sup_{x, u} W_{n, x}(u) = O(n^\beta)$. De modo que,

$$|CT_{l_n, x}^{(4)}(h) - CT_{l_n^*, x}^{(4)}(h)| = O(n^{\beta+r_1-1}) \quad c.s.,$$

que junto con **[H.12]** y **[H.13]** concluye la demostración.

Demostración del Lema 6.2. Al igual que en el Lema 6.3, la prueba de este resultado se hace exactamente igual que en el caso global, con cuidado de acotar la función de pesos por n^β en lugar de hacerlo por una constante.

Demostración del Lema 6.1. Una vez más, la prueba es una réplica de la correspondiente en la situación global, en este caso del Lema 16 de Estévez y Quintela (1999). Un esquema de la misma se muestra a continuación.

El término $CT_{l_n^*, x}^{(4)}(h)$ puede escribirse como

$$-CT_{l_n^*, x}^{(4)}(h) = \frac{1}{nn_{l_n^*}} \sum \sum_{|i-j| > l_n^*} \Gamma_x(i, j),$$

donde

$$\Gamma_x(i, j) = (H_h(X_j, X_i) - F(X_j)) D_{n,x}(X_j) - \int_a^b (H_h(y, X_i) - F(y)) D_{n,x}(y) f(y) dy$$

$$\text{y } D_{n,x}(y) = \frac{W_{n,x}(y) f(y)}{(1 - F(y))^3}. \text{ Definiendo}$$

$$\begin{aligned} \Gamma_x^*(j) &= \int (H_h(X_j, u) - F(X_j)) D_{n,x}(X_j) f(u) du - \\ &\quad - \int_a^b \int (H_h(y, u) - F(y)) D_{n,x}(y) f(y) f(u) dy du, \end{aligned}$$

$$\Psi_x(i, j) = \Gamma_x(i, j) - \Gamma_x^*(j) \quad \text{y} \quad T_x(h) = \sum \sum_{|i-j| > l_n^*} \Psi_x(i, j),$$

y observando que $n_{l_n^*} \sim n$, todo lo que tenemos que probar es

$$\sup_{h \in H_n^l} n^{-2} h^{-2k} |T_x(h)| \rightarrow 0 \quad c.s.$$

y

$$\sup_{h \in H_n^l} h \left| \sum_{j=1}^n \Gamma_x^*(j) \right| \rightarrow 0 \quad c.s..$$

La demostración de las convergencias anteriores sigue los mismos pasos que las correspondientes en el caso global (ver (22) y (23) en Estévez y Quintela (1999)), con la precaución, una vez más, de acotar las funciones de peso por n^β y posteriormente hacer uso de [H.2].

Demostración del Teorema 4.2. Se observa que

$$(26) \quad r_h(y) - r(y) = \left(\frac{f_h(y)}{1 - F_h(y)} - \frac{f(y)}{1 - F(y)} \right) \frac{1 - F_h(y)}{1 - F(y)} + \\ + \frac{(f_h(y) - r(y)\bar{F}_h(y))(\bar{F}(y) - \bar{F}_h(y))}{\bar{F}_h(y)\bar{F}(y)},$$

y puesto que $F_h(x)$ converge uniformemente a $F(x)$ sobre $x \in [a, b]$ y $h \in H_n$,

$$(27) \quad \left| \frac{(f_h(y) - r(y)\bar{F}_h(y))(\bar{F}(y) - \bar{F}_h(y))}{\bar{F}_h(y)\bar{F}(y)} \right| \leq c_n \left| \left(\frac{f_h(y)}{1 - F_h(y)} - \frac{f(y)}{1 - F(y)} \right) \frac{1 - F_h(y)}{1 - F(y)} \right|,$$

siendo c_n una sucesión de reales que tiende a cero. Por lo tanto, se obtiene que

$$ISE_x(h) = ISE_x^*(h) + o(ISE_x^*(h)),$$

y

$$ASE_x(r_h) = ASE_x^*(r_h) + o(ASE_x^*(r_h)),$$

donde

$$(28) \quad ISE_x^*(h) = \int_a^b \left[\left(\frac{f_h(y)}{1 - F_h(y)} - \frac{f(y)}{1 - F(y)} \right) \frac{1 - F_h(y)}{1 - F(y)} \right]^2 W_{n,x}(y) dy$$

y

$$(29) \quad ASE_x^*(h) = \sum_{i=1}^n \left[\left(\frac{f_h(X_i)}{1 - F_h(X_i)} - \frac{f(X_i)}{1 - F(X_i)} \right) \frac{1 - F_h(X_i)}{1 - F(X_i)} \right]^2 W_{n,x}(X_i)$$

son las versiones modificadas de los errores $ISE_x(h)$ y $ASE_x(h)$, respectivamente.

Por otra parte, cualquiera de las medidas locales modificadas puede ser vista como el error correspondiente, sin modificar, de $r_h^l(y) = \frac{f_h(y)}{\bar{F}(y)} - \frac{r(y)\bar{F}_h(y)}{\bar{F}(y)}$ considerado como

estimador de la función 0. En otras palabras, $MISE_x^*(h) = MISE_x(r'_h)$, $MSE_x^*(h) = MSE_x(r'_h)$, $ISE_x^*(h) = ISE_x(r'_h)$ y $ASE_x^*(h) = ASE_x(r'_h)$. Si a todo ello añadimos el hecho de que $r'_h(\cdot)$ puede escribirse como un estimador tipo delta, es decir,

$$r'_h(y) = n^{-1} \sum_{i=1}^n \frac{K_h(y, X_i) - r(y)(1 - H_h(y, X_i))}{1 - F(y)} = n^{-1} \sum_{i=1}^n K_h^*(y, X_i),$$

la demostración de la primera expresión en el Teorema 4.2 será análoga a la del Teorema 3.1.1 de Quintela (1992) reemplazando $K_h(\cdot)$ por $K_h^*(\cdot)$.

A continuación se da un esquema de la prueba de la segunda expresión en el Teorema 4.2. Tal y como se ha apuntado previamente, el $MISE_x^*(h)$ puede ser visto como el Error Cuadrático Medio Integrado de un estimador tipo delta de la función 0, esto es,

$$MISE_x^*(h) = E \int_a^b (r'_h(y))^2 W_{n,x}(y) dy.$$

Nótese además, que

$$(30) \quad r'_h(y) = \frac{f_h(y) - f(y)}{1 - F(y)} + \frac{(F_h(y) - F(y))r(y)}{1 - F(y)}.$$

Ahora bien, dado que es posible escribir

$$MISE_x^*(h) = E \int_a^b (r'_h(y) - Er'_h(y))^2 W_{n,x}(y) dy + \int_a^b (Er'_h(y))^2 W_{n,x}(y) dy,$$

una aplicación del Lema 4.2.4 de Estévez (2001) conduce a una descomposición del mismo en suma de dos términos, salvo una $o\left(\frac{1}{nh}\right)$. Tales términos,

$$S_x(r'_h) = \int_a^b (Er'_h(y))^2 W_{n,x}(y) dy$$

y

$$\bar{R}_x(r'_h) = \bar{E} \int_a^b (r'_h(y) - Er'_h(y))^2 W_{n,x}(y) dy$$

coinciden con sus homólogos en el caso de independencia, resultando por consiguiente de fácil manejo.

Así pues, en referencia con el término asociado al sesgo de $r'_h(\cdot)$ ($S_x(r'_h)$), es suficiente aplicar integración por partes, hacer ciertos cambios de variable y utilizar [H.9] para obtener:

$$(31) \quad \begin{aligned} Ef_h(y) - f(y) &= \int K(z) f(y - hz) dz - f(y) = \\ &= h^k \left(\frac{1}{k!} \int t^k K(t) dt \right) f^{(k)}(y) + o(h^k) \end{aligned}$$

y

$$(32) \quad \begin{aligned} EF_h(y) - F(y) &= \int K(z) F(y - hz) dz - F(y) = \\ &= h^k \left(\frac{1}{k!} \int t^k K(t) dt \right) F^{(k)}(y) + o(h^k), \end{aligned}$$

y así poder concluir que

$$(33) \quad S_x(r'_h) = h^{2k} \int_a^b (B_r(y))^2 W_{n,x}(y) dy + o(h^{2k}),$$

con

$$B_r(y) = \left(\frac{1}{k!} \int t^k K(t) dt \right) \frac{(f^{(k)}(y) + r(y) F^{(k)}(y))}{(1 - F(y))}.$$

En cuanto al término de la varianza ($\bar{R}_x(r_h)$), en virtud de (30) se tiene que:

$$\begin{aligned} \bar{R}_x(r'_h) &= \int_a^b \bar{E} \left[(f_h(y) - Ef_h(y))^2 \right] \frac{W_{n,x}(y)}{(1 - F(y))^2} dy + \\ &+ \int_a^b \bar{E} \left[(F_h(y) - EF_h(y))^2 \right] \frac{W_{n,x}(y) r^2(y)}{(1 - F(y))^2} dy + \\ &+ 2 \int_a^b \bar{E} [(f_h(y) - Ef_h(y)) (F_h(y) - EF_h(y))] \frac{W_{n,x}(y) r(y)}{(1 - F(y))^2} dy, \end{aligned}$$

reduciendo el problema de calcular varianzas y covarianzas de estimadores núcleo en el supuesto de dependencia al cálculo, notablemente más simple, de las mismas cantidades bajo independencia. Para el primer sumando es suficiente efectuar un cambio de variable y aplicar los desarrollos de Taylor de orden 0 y k a la función $f(\cdot)$ para obtener que

$$(34) \quad \begin{aligned} \overline{E} \left[(f_h(y) - Ef_h(y))^2 \right] &= n^{-1} E \left[(K_h(y, X))^2 \right] - n^{-1} [EK_h(y, X)]^2 = \\ &= \frac{1}{nh} f(y) \int K^2(t) dt + o\left(\frac{1}{nh}\right). \end{aligned}$$

Análogamente, para el sumando que involucra a la función de distribución y su suavizador, se consigue:

$$(35) \quad \begin{aligned} \overline{E} \left[(F_h(y) - EF_h(y))^2 \right] &= n^{-1} E \left[(H_h(y, X))^2 \right] - n^{-1} [EH_h(y, X)]^2 = \\ &= \frac{F(y)(1-F(y))}{n} - \frac{h}{n} \left(\int 2zK(z)H(z)dz \right) f(y) + o\left(\frac{h}{n}\right) = \\ &= o\left(\frac{1}{nh}\right). \end{aligned}$$

Finalmente, el término de la covarianza se puede acotar mediante:

$$\begin{aligned} \overline{E} \left[(f_h(y) - Ef_h(y)) (F_h(y) - EF_h(y)) \right] &= \\ &= \frac{1}{n} E [K_h(y, X) H_h(y, X)] - \frac{1}{n} EK_h(y, X) EH_h(y, X) \leq C \frac{1}{n} = o\left(\frac{1}{nh}\right). \end{aligned}$$

Por lo tanto, se obtiene que

$$(36) \quad \overline{R}_x(r_h) = \frac{1}{nh} \int_a^b V_r^2(y) W_{n,x}(y) dy + o\left(\frac{1}{nh}\right),$$

con

$$V_r^2(y) = \left(\int K^2(t) dt \right) \frac{f(y)}{(1 - F(y))^2}.$$

En definitiva, combinando (33) y (36) se consigue el resultado deseado.

7. AGRADECIMIENTOS

Deseamos expresar nuestra gratitud al profesor Philippe Vieu (Universidad Paul Sabatier, Toulouse, Francia) por sus comentarios, que nos han ayudado a realizar una mejor presentación del trabajo. También deseamos agradecer la gentileza del profesor Angel Venedikov (Academia de Ciencias de Bulgaria), que nos cedió un amplio fichero de datos con los terremotos sucedidos en California durante el periodo de tiempo estudiado en este artículo.

Este trabajo ha sido financiado por la Xunta de Galicia mediante el proyecto de investigación XUGA 10503A98 y por la DGES bajo el proyecto de investigación PB 98-0182-C02-01.

8. REFERENCIAS

- Ahmad, I. A. (1976). «Uniform strong convergence of the generalized failure rate estimate». *Bull. Math. Statist.*, 17, 77-84.
- Bowman, A. (1984). «An alternative method of cross-validation for the smoothing of density estimates». *Biometrika*, 71, 353-360.
- Cox, D. R. & Oakes, D. (1984). *Analysis of survival data*. Chapman and Hall.
- Doukhan, P. (1994). «Mixing: properties and examples». *Lecture Notes in Statistics*, 85, Springer-Verlag.
- Chu, C. K. & Marron, J. S. (1991). «Comparison of two bandwidth selectors with dependent errors». *The Annals of Statistics*, 4, 1906-1918.
- Estévez, G. (2001). *Estimación tipo núcleo de la función de razón de fallo bajo condiciones de dependencia*. Tesis Doctoral. Universidad de Santiago de Compostela.
- Estévez, G. & Quintela, A. (1999). «Nonparametric estimation of the hazard function under dependent conditions». *Communications in Statistics-Theory and Methods*, 28, 10, 2297-2331.
- Gasser, T. & Müller, H. G. (1984). «Estimating regression functions and their derivatives by the kernel method». *Scandinavian Journal of Statistics*, 11, 171-185.

- Györfi, L., Härdle, W., Sarda, P. & Vieu, P. (1990). «Nonparametric curve estimation from time series», *Lecture Notes in Statistics*, 60, Springer-Verlag.
- Härdle, W. & Vieu, P. (1992). «Kernel regression smoothing of time series», *Journal of Time Series Analysis*, 13, 209-232.
- Hart, J. & Vieu, P. (1990). «Data-driven bandwidth choice for density estimation based on dependent data». *The Annals of Statistics*, 18, 873-890.
- Hassani, S., Sarda, P. & Vieu, P. (1986). «Approche non paramétrique en théorie de la fiabilité». *Revue de Statistiques Appliquées*, vol. XXXV, n.º 4.
- Hollander, M. & Proschan, F. (1984). «Nonparametric concepts and methods in reliability». *Handbook of Statistics*, vol. 4. P. R. Krishnaiah and P. K. Sen, eds., pp. 613-655, North-Holland, Amsterdam.
- Klein, J. P. & Moeschberger, M. L. (1997). «Survival Analysis». *Techniques for censored and truncated data*. Springer-Verlag.
- Lee, E. T. (1992). *Statistical methods for survival data analysis*. Wiley Series in Probability and Mathematical Statistics.
- Lomnitz, C. & Hax, A. (1967). «Clustering in aftershock sequences». *The earth beneath the continents. Geophys. Monograph. 10, Am. Geophys. Union*.
- Prakasa Rao, B. L. S. & Van Ryzin, J. (1985). «Asymptotic theory for two estimators of the generalized failure rate». *Statistical Theory and Data Analysis*. K. Matusita, ed., North-Holland, Amsterdam.
- Quintela, A. (1992). *Cálculo del parámetro de suavización en la estimación no paramétrica de curvas con datos dependientes*. Tesis Doctoral, Universidad de Santiago de Compostela.
- Quintela, A. (1994). «A plug-in technique in nonparametric regression with dependence». *Communications in Statistics-Theory and Methods*, 23, 2581-2603.
- Quintela, A. & Vilar, J. M. (1992). «A local cross-validation algorithm for dependent data». *TEST*, 1, 123-153.
- Reiss, R. D. (1981). «Nonparametric estimation of smooth distribution functions». *Scandinavian Journal of Statistics*, 8, 116-119.
- Rice, J. & Rosenblatt, M. (1976). «Estimation of the log survivor function and hazard function». *Sankhyā*, A, 38, 60-78.
- Rudemo, M. (1982). «Empirical choice of histograms and kernel density estimates». *Scandinavian Journal of Statistics*, 9, 65-78.
- Sarda, P. & Vieu, P. (1989). «Estimation non paramétrique de la fonction de hasard». *Cahiers du C.E.R.O.*, 31, 241-265.
- Singpurwalla, N. D. & Wong, M. (1983). «Estimation of the failure rate-A survey of nonparametric methods. Part I: Non-Bayesian methods». *Communications in Statistics-Theory and Methods*, 12, 559-588.

- Udias, A. & Rice, J. (1975). «Statistical analysis of microearthquakes activity near San Andreas Geophysical Observatory. Hollister, California». *Bulletin of the Seismological Society of America*, 65, 809-828.
- Vere-Jones, D. (1970). «Stochastic models for earthquake occurrence». *Journal of the Royal Statistical Society, Ser. B*, 32, 1-62.
- Vere-Jones, D. & Davies, D. (1966). «A statistical survey of earthquakes in the main seismic region of New Zealand, II. Time series analysis». *New Zealand J. Geol. Geophys*, 9, 251-284.
- Vieu, P. (1991a). «Quadratic errors for nonparametric estimates under dependence». *Journal of multivariate analysis*, 39, 324-347.
- Vieu, P. (1991b). «Nonparametric regression: optimal local bandwidth choice». *Journal of the Royal Statistical Society, Ser. B*, 53, 453-464.
- Watson, G. S. & Leadbetter, M. R. (1964a). «Hazard Analysis I». *Biometrika*, 51, 175-184.
- Watson, G. S. & Leadbetter, M. R. (1964b). «Hazard Analysis II». *Sankhyā, A*, 26, 110-116.
- Youndjé, É., Sarda, P. & Vieu, P. (1996). «Optimal smooth hazard estimates». *TEST*, 5, 379-394.

ENGLISH SUMMARY

NONPARAMETRIC ESTIMATION OF THE HAZARD FUNCTION: APPLICATIONS TO SEISMOLOGY

GRACIELA ESTÉVEZ PÉREZ
ALEJANDRO QUINTELA DEL RÍO
Universidad de A Coruña*

We study the nonparametric estimation of the risk or hazard function of a real random variable. Through an uncensored random sample of data X_1, X_2, \dots, X_n , not necessarily independent, we consider an estimator built as a ratio of a kernel estimator of the density function and a kernel estimator of the survival function. We consider the bandwidth selection problem in this estimator.

We show, by means of a simulation example, the improvement in the estimates if we use this estimator instead of the one that uses in the denominator the empirical distribution function. Finally, we present a practical application to data of earthquakes in California (United States) and Granada (Spain).

Keywords: Nonparametric estimation, strong mixing processes, cross-validation, hazard function

AMS Classification (MSC 2000): 62G05, 62G20, 62M99

*Departamento de Matemáticas-Facultad de Informática. Universidad de A Coruña. Campus de Elviña, s/n. 15071 A Coruña. Teléfono: 981167000, ext.: 2109.

–Received July 1999.

–Accepted October 2001.

1. INTRODUCTION

Let X be a real random variable with probability density function $f(\cdot)$ and distribution function $F(\cdot)$ with respect to some σ -finite measure on \mathbb{R} . The **failure rate** or **hazard function** is defined by $r(x) = \lim_{\Delta_x \rightarrow 0} \frac{P(x \leq X < x + \Delta_x / X \geq x)}{\Delta_x}$.

By the definition of conditional probability, we have that

$$(1) \quad r(x) = \frac{f(x)}{1 - F(x)} = \frac{f(x)}{\bar{F}(x)},$$

considered when $\bar{F}(x) > 0$, that is, $r(\cdot)$ is defined in the set $S = \{x \in \mathbb{R} / \bar{F}(x) > 0\}$.

If we consider that the random variable X measures the «failure time», or the time of occurrence of a particular event in an homogeneous population, $r(x)\Delta_x$ can be interpreted as the approximate probability of one subject «fails» in the time interval $[x, x + \Delta_x)$, given the subject has survived to time x , i.e. the instantaneous probability of failure at x , given the last fail happened in the moment 0.

Through this point of view, the estimation of the hazard function is a problem of considerable interest, specially to medical researchers, logistics planners, reliability engineers, and seismologists. In medical research, there exists a lot of applications of the hazard estimation (Hassani, Sarda and Vieu (1986)). In seismology, the hazard rate function might be thought as the instantaneous risk of the occurrence of an earthquake at time x , known that at time 0 the last earthquake has happened (Rice and Rosenblatt (1976)).

A practical question is the estimation of $r(\cdot)$ based on a random sample X_1, \dots, X_n of X . This problem has been extensively discussed in the literature in the case of i.i.d. random variables. See, for example, Watson and Leadbetter (1964a, b), Ahmad (1976), Hollander and Proschan (1984) and Hassani, Sarda and Vieu (1986).

However, in several fields of applications the independence assumption is far from being realistic. This is for instance the case of microearthquake studies (Rice and Rosenblatt (1976)). Therefore, we will consider that the random data pertain to a strictly stationary sequence of strong mixing variables (Rosenblatt (1956)).

In this paper, we approach the nonparametric estimation of hazard function (1) under dependence conditions and no censored. In the next section we define the method of estimation and recall some of its properties. Section 3 shows some simulations that exhibit such properties. In Section 4 we study the local cross-validation method for the calculation of the bandwidths. Finally, in Section 5 we present an application to a set of microearthquake data.

2. KERNEL ESTIMATION OF HAZARD FUNCTION UNDER DEPENDENCE CONDITIONS

Several estimators proposed, and in particular the one which will be object of study in this work, are built as the ratio of a nonparametric estimator of $f(\cdot)$ and a nonparametric estimator of $\bar{F}(\cdot) = 1 - F(\cdot)$. (For a survey on different types of estimators of the hazard function see Rice and Rosenblatt (1976)). We will consider a smooth estimator of $r(\cdot)$, introduced by Watson and Leadbetter (1964a, b), but we will work in a context of α -mixing data (Rosenblatt (1956)):

$$(2) \quad r_h(x) = \frac{f_h(x)}{1 - F_h(x)},$$

where $f_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$ is the known Parzen-Rosenblatt estimator of $f(\cdot)$, and $F_h(\cdot)$ is the kernel estimator of $F(\cdot)$ defined by

$$(3) \quad F_h(x) = \frac{1}{n} \sum_{i=1}^n H\left(\frac{x - X_i}{h}\right) = \int_{-\infty}^x f_h(t) dt.$$

with $K(\cdot)$ a kernel function, $H(x) = \int_{-\infty}^x K(u) du$ and $h = h(n) \in \mathbb{R}^+$ is the smoothing parameter, or bandwidth.

This estimator presents some advantages as compared to estimators that use the empirical distribution function to estimate the function $F(\cdot)$ $\left(r_{h,1}(x) = \frac{f_h(x)}{1 - F_n(x)}\right)$, such as its continuity, and presents improvements considered in terms of quadratic errors (see e.g. Youndjé, Sarda and Vieu (1996) and references therein). Consistency properties and asymptotic normality were studied by Estévez and Quintela (1999). The authors have also developed a smoothing parameter selection method for this estimator, through the least-squares cross-validation, showing that it is asymptotically optimal, in the sense that

$$\frac{d_I(\hat{h}(l_n))}{\inf_h d_I(h)} \rightarrow 1 \text{ a.s.}$$

where $d_I(\cdot)$ is any of quadratic errors $ISE(\cdot)$, $ASE(\cdot)$ or $MISE^*(\cdot)$ and $\hat{h}(l_n)$ is the cross-validation bandwidth with l_n a parameter depending of the amount of dependence.

3. SIMULATIONS

In order to compare the estimators $r_h(\cdot)$ and $r_{h,1}(\cdot)$ we carried out the following study of simulation: we generated M samples with N data from an autoregressive process of order one with distribution $N(0, 1)$ or from a m -dependent model with distribution $\Gamma(1, 1.5)$. For each one of the samples we calculated the cross-validation and the optimal bandwidth and the corresponding errors. The obtained results assure the kernel estimator ($r_h(\cdot)$) is behaved better than $r_{h,1}(\cdot)$. With regard to bandwidth selection, the results do not agree with what happened in density estimation (Hart and Vieu (1990)) or regression estimation (Quintela (1994)). Here, we can observe that the cross-validation bandwidth practically does not change with l_n . The horizontal convergence rates of cross-validation procedure maybe could explain this small influence of l_n .

4. LOCAL BANDWIDTH CHOICE

In this section we investigate a data-driven local selection method pertaining to the minimization, in each point x of an interval $[a, b]$, of the pointwise Mean Squared Error

$$(4) \quad MSE_x(h) = E(r_h(x) - r(x))^2.$$

To select, in each point x , a data-driven local bandwidth h_x we propose to minimize some estimate of MSE_x . The criterion used to estimate MSE_x is a local adaptation of the usual cross-validation rule. In effect, for each point x in the interval $[a, b]$, we will chose the bandwidth $\hat{h}_x(l_n)$ minimizing the function:

$$(5) \quad CV_{x,l_n}(h) = \int_a^b r_h^2(y) W_{n,x}(y) dy - 2n^{-1} \sum_{i=1}^n \frac{f_h^{-i}(X_i)}{(1 - F_h^{-i}(X_i))(1 - F_n(X_i))} W_{n,x}(X_i),$$

where $f_h^{-i}(\cdot)$ and $F_h^{-i}(\cdot)$ are defined as in the global case and $W_{n,x}(\cdot)$ is a sequence of weight functions concentrated around x .

The local bandwidth also is asymptotically optimal, in the sense that

$$\sup_{x \in [a,b]} \frac{d_x(\hat{h}_x(l_n))}{\inf_{h \in H_n} d_x(h)} \rightarrow 1 \quad a.s.,$$

where $d_x(\cdot)$ is any of local quadratic errors $ASE_x(\cdot)$, $ISE_x(\cdot)$ or $MISE_x^*(\cdot)$. It is important to note that the previous result uses the asymptotic equivalence between such errors, that is,

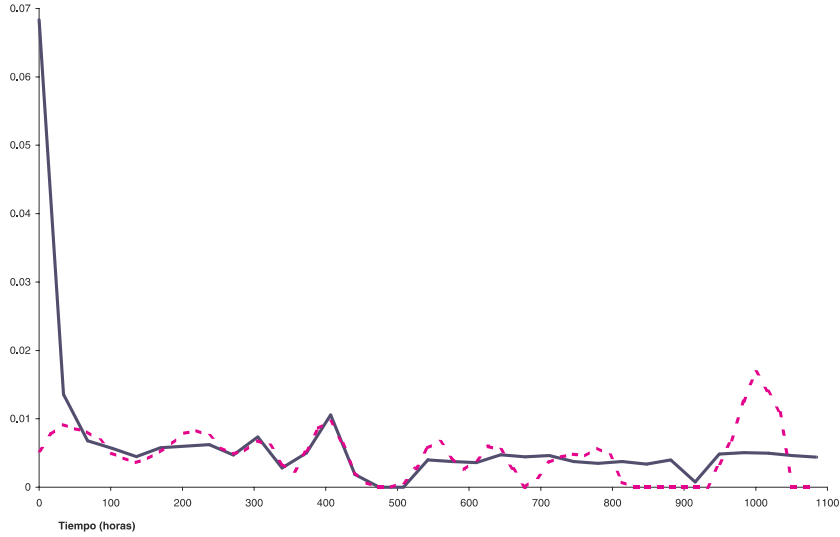
$$\sup_{h \in H_n} \frac{|d_x(h) - d'_x(h)|}{d'_x(h)} \rightarrow 0 \quad a.s..$$

These results are collected in the Theorems 4.1 and 4.2 of the Spanish version of the paper.

5. APPLICATIONS TO SEISMOLOGY

Our applications deal with the seismicity analysis. We have applied the cross-validation procedure and the hazard kernel estimation to a geological set of data. Our time series are obtained of the global set of earthquakes observations recorded in two well different regions: California (USA) and Granada (Spain). From the whole data registered, we select our data sets and we consider the time series formed by the time intervals (in hours) between the occurrence of each two earthquakes. These samples are composed of dependent data (effecting several tests of independence, in no case it could be admitted the independence hypothesis between the data). Applying the cross-validation procedure to these time series, we obtain, for the region of Granada, the hazard kernel estimate using the cross-validation bandwidth.

We have also obtained the local bandwidth and the corresponding estimations. The resulting curves have been represented in the following figure:



We can observe that the local method allows us to estimate more efficiently the hazard function over the full sample interval than using a global estimate.

6. ACKNOWLEDGMENTS

We would like to express our gratitude to the professor Philippe Vieu (University Paul Sabatier, Toulouse, France) for its comments, which have helped us to prepare a better version of the paper. Thanks are also due to the professor Angel Venedikov (Academic of Sciences of Bulgaria), that gave us a large data file with the earthquakes of California. This research was financed by the Xunta de Galicia under project XUGA 10503A98 and by DGES under research project PB 98-0182-C02-01.

7. REFERENCES

- Ahmad, I. A. (1976). «Uniform strong convergence of the generalized failure rate estimate». *Bull. Math. Statist.*, 17, 77-84.
- Estévez, G. & Quintela, A. (1999). «Nonparametric estimation of the hazard function under dependent conditions». *Communications in Statistics-Theory and Methods*, 28, 10, 2297-2331.
- Hart, J. & Vieu, P. (1990). «Data-driven bandwidth choice for density estimation based on dependent data». *The Annals of Statistics*, 18, 873-890.
- Hassani, S., Sarda, P. & Vieu, P. (1986). «Approche non paramétrique en théorie de la fiabilité». *Revue de Statistiques Appliquées*, vol. XXXV, n.º 4.
- Hollander, M. & Proschan, F. (1984). «Nonparametric concepts and methods in reliability». *Handbook of Statistics*, vol. 4. P. R. Krishnaiah and P. K. Sen, eds., pp. 613-655, North-Holland, Amsterdam.
- Quintela, A. (1994). «A plug-in technique in nonparametric regression with dependence». *Communications in Statistics-Theory and Methods*, 23, 2581-2603.
- Rice, J. & Rosenblatt, M. (1976). «Estimation of the log survivor function and hazard function». *Sankhyā*, A, 38, 60-78.
- Rosenblatt, M. (1956). «A central limit theorem and a strong mixing condition». *Proceedings of the National Academic Science*, 42, 43-47.
- Watson, G. S. & Leadbetter, M. R. (1964a). «Hazard Analysis I». *Biometrika*, 51, 175-184.
- Watson, G. S. & Leadbetter, M. R. (1964b). «Hazard Analysis II». *Sankhyā*, A, 26, 110-116.
- Youndjé, É., Sarda, P. & Vieu, P. (1996). «Optimal smooth hazard estimates». *TEST*, 5, 379-394.